Proceedings of the SPIE, vol. 3461, pp. 91-98, Advanced Signal Processing Algorithms, Architectures and Implementations, San Diego, CA, October 1998.

Data-Driven Optimization of Time and Frequency Resolution for Radar Transmitter Identification

Bradford W. Gillespie and Les E. Atlas
Interactive Systems Design Laboratory, Department of Electrical Engineering
University of Washington
Box 352500
Seattle, Washington 98195-2500 USA

ABSTRACT

An entirely new set of criteria for the design of kernels (*i.e.* generating functions) for time-frequency representations (TFRs) has been recently proposed.[1,2,3] The goal of these criteria is to produce kernels (and thus, TFRs) which will enable accurate classification without explicitly defining, *a priori*, the underlying features that differentiate individual classes. These kernels, which are optimized to discriminate among multiple classes of signals, are referred to as *signal class-dependent kernels*, or simply *class-dependent kernels*. Here this technique is applied to the problem of radar transmitter identification. Several modifications to our earlier approach have been incorporated into the processing, and are detailed here. It will be shown that an overall classification rate of 100% can be achieved using our new augmented approach, provided exact time registration of the data is available. In practice, time registration can not be guaranteed. Therefore, the robustness of our technique to data misalignment is also investigated. A measurable performance loss is incurred in this case. A method for mitigating this loss by incorporating our class-dependent methodology within the framework of classification trees is proposed.

Keywords: Time-Frequency, Classification, Detection, Radar, Transmitter Identification, Unintentional Modulation

1. INTRODUCTION

One application of interest in radar signal processing is the detection and classification of individual radar signals. The goal is to identify the particular transmitter from which the signal originated. Individual transmitter identification can be accomplished by exploiting the unintentional modulation present in these radar signals. This modulation is a result of subtle variations between particular transmitter components, and acts as a signature for an individual radar station.

A variety of techniques could be used to identify individual transmitters using the unintentional modulation present on the radar signal. Instead of imposing a time, frequency or combined time-frequency approach, it may be potentially more informative to allow the classifier to determine what information is needed to accurately separate the data. Recently, we have devised a method that allows the classification task to, given adequate and representative training data, ascertain the relative role of time and frequency resolution in classification. For example, if the center frequency were important for transmitter identification, it would be best to have high resolution in frequency and little or no resolution in time. This optimal smoothing in time and frequency is determined automatically from the training data.

Our approach is based on the premise that automatic detection and classification systems should be provided with only enough input resolution to achieve needed performance. Namely, resolution that is too great will potentially require a large

Email Correspondence: B. W. Gillespie brad@ee.washington.edu
L. E. Atlas atlas@ee.washington.edu

Report Documentation Page	*Form Approved OMB No. 0704-0188*

1. REPORT DATE **1998**	2. REPORT TYPE	3. DATES COVERED **-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Data-Driven Optimization of Time and Frequency Resolution for Radar Transmitter Identification	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Office of Naval Research,875 North Randolph Streeet,Arlington,VA,22203-1995**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
An entirely new set of criteria for the design of kernels (i.e. generating functions) for time-frequency representations (TFRs) has been recently proposed.1,2,3 The goal of these criteria is to produce kernels (and thus, TFRs) which will enable accurate classification without explicitly defining, a priori, the underlying features that differentiate individual classes. These kernels which are optimized to discriminate among multiple classes of signals, are referred to as signal class-dependent kernels, or simply class-dependent kernels. Here this technique is applied to the problem of radar transmitter identification. Several modifications to our earlier approach have been incorporated into the processing, and are detailed here. It will be shown that an overall classification rate of 100% can be achieved using our new augmented approach, provided exact time registration of the data is available. In practice, time registration can not be guaranteed. Therefore, the robustness of our technique to data misalignment is also investigated. A measurable performance loss is incurred in this case. A method for mitigating this loss by incorporating our class-dependent methodology within the framework of classification trees is proposed.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**8**	

detector or classifier training set and will be sensitive to irrelevant features and/or noise. Large dimensionality detectors and classifiers are also computationally expensive and slow. It should be noted that we are not referring to or bound by implicit Heisenburg or window-related resolution limitations — we are instead explicitly limiting the resolution to optimize the accurate identification of radar transmitters.

2. BACKGROUND

Modern time-frequency representation (TFR) research often begins by selecting a kernel (*i.e.* generating function) $\Phi[n,\tau]$ that operates upon an instantaneous autocorrelation function:

$$R[n,\tau] = \sum_{n'=n-N}^{n+N} x[n']x[n'+\tau] \,. \tag{1}$$

The resultant TFR, $P[n,k]$, arises from the discrete Fourier transform (in τ) of the results of multiplying the kernel (in τ) and convolving the kernel (in n) with the instantaneous autocorrelation function, $R[n,\tau]$. As an alternative, a discrete Fourier transform (in n) can be applied to the instantaneous autocorrelation function, $R[n,\tau]$, to yield an ambiguity function:

$$A[\eta,\tau] = \boldsymbol{F}_n\{R[n,\tau]\} = \sum_{n=0}^{M-1} R[n,\tau]\,\mathrm{e}^{-j\frac{2\pi}{M}n\eta} \,. \tag{2}$$

There is an equivalent kernel, $\phi[\eta,\tau]$, which operates multiplicatively in both dimensions upon the ambiguity function, $A[\eta,\tau]$. These two kernels are also related by a discrete Fourier transform (in n):

$$\phi[\eta,\tau] = \boldsymbol{F}_n\{\Phi[n,\tau]\} = \sum_{n=0}^{M-1} \Phi[n,\tau]\,\mathrm{e}^{-j\frac{2\pi}{M}n\eta} \,. \tag{3}$$

Any non-zero extent of $\phi[\eta,\tau]$ in η and/or τ can effect a smoothing on $P[n,k]$ in time and/or frequency respectively. For example, if $\phi[\eta,\tau] = 0$ for all values except those on the $\eta = 0$ axis, then all temporal information is smoothed and only steady-state frequency information is retained in $P[n,k]$. In past time-frequency research, kernels for quite a number of properties, such as finite-time support and minimizing quadratic interference, have been derived. Although some of these representations may offer advantages in classification of certain types of signals, the goal of sensitive detection or accurate classification has not been explicit. The ability of the aforementioned kernel to reduce time and/or frequency resolution, embodied within the explicit goal of optimal classification (*i.e.* minimum number of classification errors), is the basis for the approach outlined below. When the kernel, $\phi[\eta,\tau]$, is designed with the goal of optimal classification we refer to it as the *signal class-dependent kernel*, or simply *class-dependent kernel*. Furthermore, we refer to the corresponding TFR, $CD[n,k]$, as the *class-dependent TFR*.

3. OUR APPROACH AND METHODS

Data provided by the Naval Research Lab (NRL) was utilized in this work.[4] This data set contains ten radar pulses from four transmitters. This data comprises three tests from each of the four sources called A2, CCC2, F2, and H2. These will be denoted as class one through four. Each pulse contains 180 complex samples (*i.e.* in-phase and quadrature components). An example of a radar pulse from each class is shown in Figure 1, and the corresponding TFR in Figure 2.

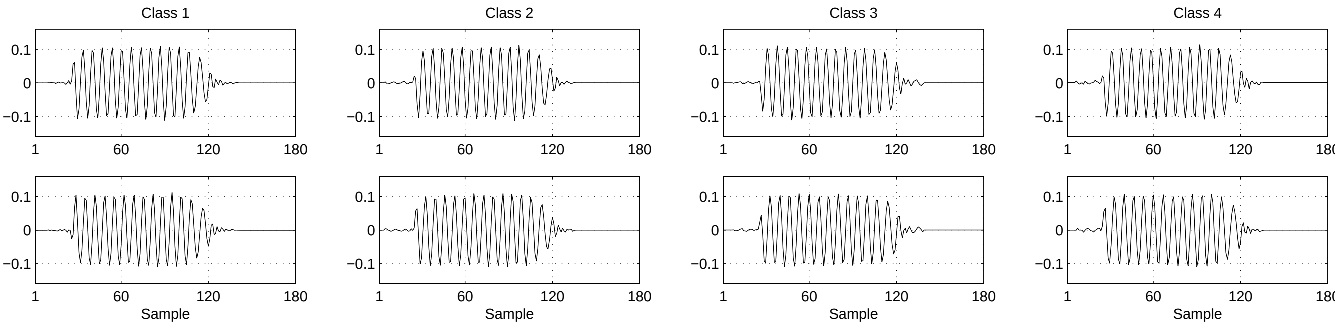

Figure 1. Example signal from each class, in-phase (top) and quadrature (bottom).

In order to experimentally study a proposed detection system the data was randomly divided into nine training examples and one test example for each of the four classes, and training and testing performed. This process was repeated 10,000 times, and the performance averaged, to yield an accurate performance estimate of the system.

Our approach is a modification of the signal class-dependent method that has been described in more detail before.[1,2,3] The previously described approach finds the single kernel, $\phi[\eta,\tau]$, which maximizes the distance, in a mean-square sense, between the estimated ambiguity functions for each of C different classes. Defining a kernel matrix as $\boldsymbol{\phi} = \phi[\eta,\tau]$ and an ambiguity matrix for class c as $\boldsymbol{A}_c = A_c[\eta,\tau]$, the kernel is selected to satisfy:

$$\arg\max_{\boldsymbol{\phi}} \left\{ \sum_{c'=1}^{C} \sum_{c''=c'+1}^{C} \left\| \boldsymbol{\phi} \circ \boldsymbol{A}_{c'} - \boldsymbol{\phi} \circ \boldsymbol{A}_{c''} \right\|_2^2 \right\}, \tag{4}$$

where $\circ$ represents the Hadamard product (*i.e.* an element-by-element product).

In practice, this maximization is accomplished by rank-ordering the kernel points according to separation between classes. Choosing the kernel point with the largest between-class separation corresponds to a maximum separation between classes. Thus for actual classification of an unknown time series, the ambiguity function is multiplied, in η and τ, by a binary kernel mask, which is set to "1" at one optimal and, optionally, subsequently lower-ranked kernel points (often required in practice).❶ These kernel points, depending on their locations, effect a smoothing in time and/or frequency of the unknown

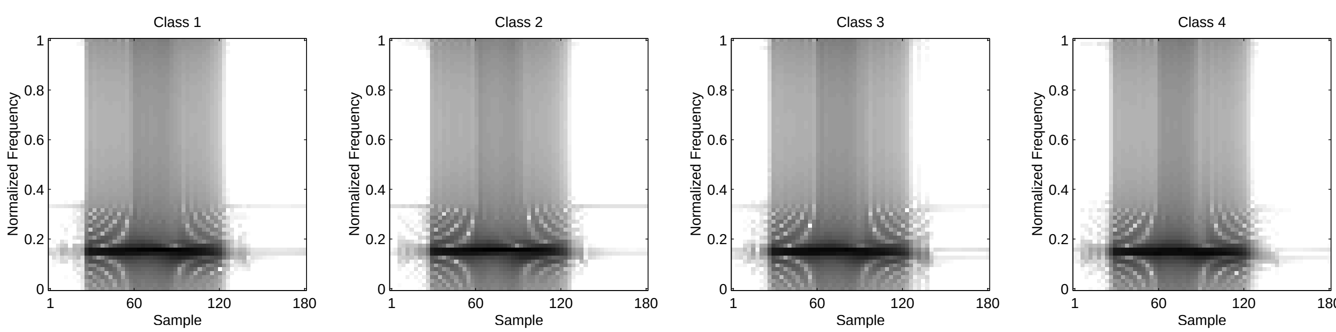

Figure 2. Log magnitude of the TFR ($20\log_{10}|\boldsymbol{P}|$) corresponding to the signals shown in Figure 1. The largest magnitude is represented by the darkest gray-scale value.

---

❶ This binary mask is selecting, in effect, "features" from the set of points that make up the ambiguity function.

data. The smoothed version is then compared to a smoothed representative from each class, derived during training. As an added benefit, the class-dependent ambiguity function $(\phi \circ \boldsymbol{A}_c)$ can be transformed into a class-dependent time-frequency representation $CD_c[n,k]$. The implicit optimal time-frequency smoothing can then be viewed.

As we have recently found, the above mean-square distance is inadequate to handle the wide range of within-class variance seen in real-world applications. Thus, we have modified our earlier approach to find the kernel, ϕ, that optimizes a Fisher's discriminant distance given by:[5]

$$\mathrm{FDR}[\eta,\tau] = \frac{\sum_{c'=1}^{C}\sum_{c''=c'+1}^{C}\left(\mu_{c'}^{\eta,\tau}-\mu_{c''}^{\eta,\tau}\right)^2}{\sum_{c=1}^{C}\left(\sigma_c^{\eta,\tau}\right)^2} = \frac{\sum_{c'=1}^{C}\sum_{c''=c'+1}^{C}\left\{\frac{1}{I}\sum_{i=1}^{I}A_{c'}^{i}[\eta,\tau]-\frac{1}{I}\sum_{i=1}^{I}A_{c''}^{i}[\eta,\tau]\right\}^2}{\sum_{c=1}^{C}\left\{\left[\frac{1}{I}\sum_{i=1}^{I}\left|A_c^i[\eta,\tau]\right|^2-\left|\frac{1}{I}\sum_{i=1}^{I}A_c^i[\eta,\tau]\right|^2\right]\right\}}, \tag{5}$$

where $A_c^i[\eta,\tau]$ is an element from the ambiguity function of the i^{th} training example from class c; $\mu_c^{\eta,\tau}$ and $\sigma_c^{\eta,\tau}$ are the estimated mean and standard deviation of the I training examples of $A_c[\eta,\tau]$. The Fisher's discriminant distance provides a rank-ordering of kernel points for classification. The optimal number of points is determined by evaluating the classifier performance using the K best kernel points (*i.e.* the K points with the largest Fisher's discriminant distance). K_{opt} is selected to be the K for which the probability of correct classification is greatest.

To classify a particular unknown test signal, an M by M point ambiguity function is estimated from the signal. After masking with the previously determined kernel, the class of the unknown signal is estimated via a maximum likelihood (ML) detector.[6] The mean and covariance statistics of the selected K points for each class (utilized by the ML detector) are estimated from the training data.

PREPROCESSING

Before classification, each data segment is preprocessed. First, each radar pulse is individually demeaned and then normalized to a standard deviation of one, in order to prevent classification based on irrelevant or variable features. The selection of the second step in preprocessing is more involved. Its necessity is an outgrowth of the particular classification technique employed. The center frequency of the transmitter is a variable parameter. Because our method seeks to find a time-frequency representation that maximizes between-class separation, if a particular class in the training set contains a center frequency bias, this will be used as an essential class discriminator. The variability of this parameter makes this unusable as a discriminatory feature. There are three possible solutions to ensure this feature is not incorporated into the classifier.

1. Given enough representative data from each transmitter (presumably including variability in the center frequency) the classifier will discard this feature as a possible means of classification. This is equivalent to the classifier "learning" that center frequency is irrelevant.

2. Only the magnitude of the radar pulse is used for classification. This presumes that there is enough information in the envelope of the radar signature to discriminate classes.

3. The data set is preprocessed to modulate all pulses to the same center frequency. This involves estimation of the center frequency of each pulse and modulation to a new predetermined center frequency.

Due to the size of the data set provided, the latter method is preferred. The large signal to noise ratio of this data makes estimation of the center frequency of the signal relatively easy. It was determined that 34 out of the 40 examples had a center frequency of 0.151(2π) radians per second while 6 pulses (all from class one) had a center frequency of 0.145(2π) radians per second.

The selected preprocessing algorithm for this data was to modulate all signals to a center frequency of 0.151(2π) radians per second. Once this preprocessing algorithm is applied to the data, transmitter identification is implemented as described above.

4. RESULTS

In the provided data set, all examples are time-aligned precisely. In practice, exact time alignment of the training and testing data can not be assured. Therefore, robustness of this technique to timing jitter is investigated. Jitter was introduced by randomly varying the start of each data segment, and testing the retrained classifier on this "new" data set. The three cases that will be presented here are:

Case 1. classification using the original data which contains precise time alignment between all examples,

Case 2. classification using the original data with timing jitter uniformly distributed over the interval ± one sample, and

Case 3. classification using the original data with timing jitter uniformly distributed over the interval ± five samples.

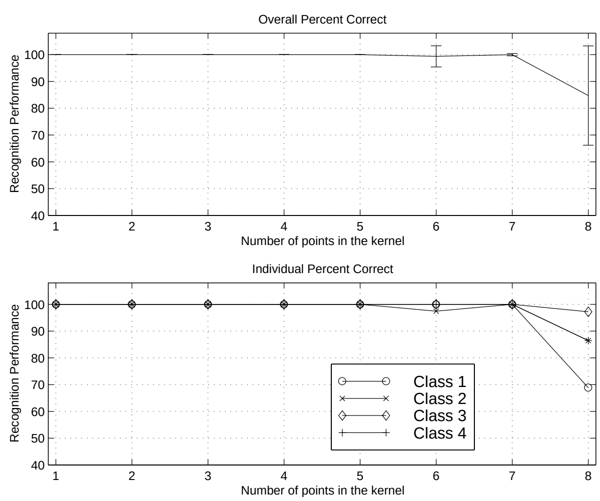

Figure 3. Rank-order curves for Case 1. Overall percent correct classification (top), bars represent ± one standard deviation. Individual percent correct classification for each class (bottom).

CASE 1.

In the original data set all the radar pulses are perfectly time aligned with respect to the envelope of the signal. Under these conditions (using a 64 by 64 point ambiguity function), 100% correct classification was achieved. To generate this estimate, the classifier was trained and tested 10,000 times and the results averaged to yield an overall measure of performance. The rank order curve for this case is shown in Figure 3. If the top 1 to 5 kernel points are used the classifier is able to discriminate all four classes perfectly. When the classifier uses more than four kernel points, the overall classification rate drops to a low of 84%.

By inverting the resulting class-dependent kernel, the class-dependent TFR can be viewed. An example from each class is given in Figure 4. This technique has determined that a combination of time and frequency information is important for classification.

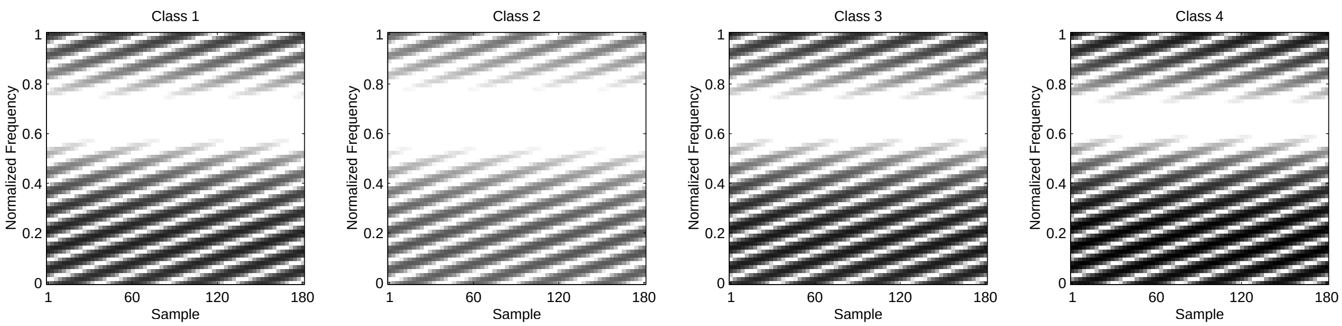

Figure 4. Example of the log magnitude of the class-dependent TFR ($20\log_{10}|\boldsymbol{CD}|$) from each class under Case 1. These are generated using four kernel points. The largest magnitude is represented by the darkest gray-scale value.

CASE 2.

Timing jitter uniformly distributed over the interval of ± 1 sample was introduced into the data. Under this condition, a 97% overall correct classification rate was achieved when the optimal number of kernel points was used. The rank order curve is shown in Figure 5 for K = 1 through K = 8. The optimal kernel for classification contains only points on the $\eta = 0$ axis. This corresponds to retaining only frequency information in the class-dependent TFR. An example class-dependent TFR from each class is shown in Figure 6.

CASE 3.

Timing jitter uniformly distributed over the interval of ± 5 samples was introduced into the data. This was intended to mimic a condition of severe input timing jitter. Under this situation, the classifier performed the same as in minor time jitter case (Case 2). We conclude that our method is insensitive to the amount of time alignment variability in the input data, assuming exact time alignment can not be assured. However, there is a significant performance gain if exact time alignment can be assured, as shown in Case 1. This gain is attributed to the classifier being able to exploit time information in classification as evidenced by comparing Figure 4 to Figure 6.

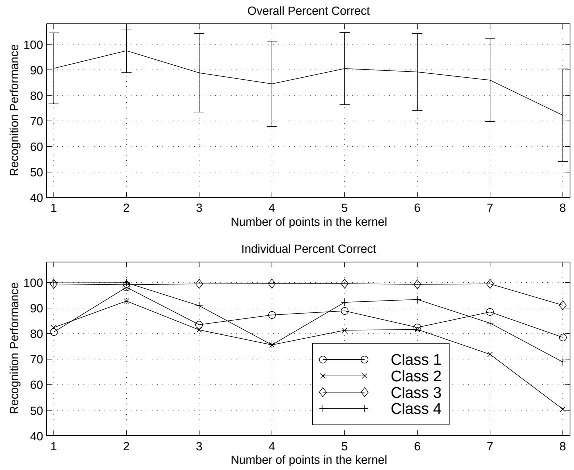

Figure 5. Rank-order curves for Case 2. Overall percent correct classification (top), bars represent ± one standard deviation. Individual percent correct classification for each class (bottom).

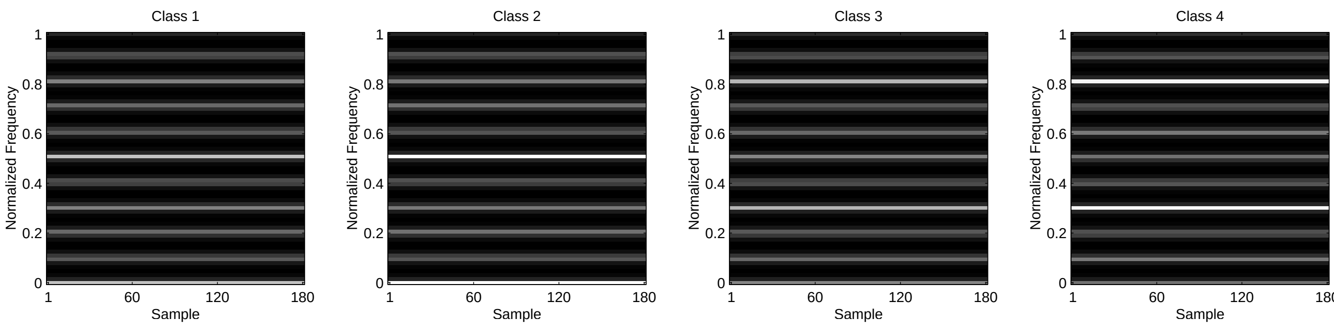

Figure 6. Example of the log magnitude of the class-dependent TFR ($20\log_{10}|\boldsymbol{CD}|$) from each class under Case 2. These are generated using two kernel points. The largest magnitude is represented by the darkest gray-scale value.

5. ONGOING WORK

Currently, to mitigate the effect of data jitter on the classification rate, a combination of class-dependent classification and classification trees[7] is being investigated. This method involves cascading multiple class-dependent classifiers to discriminate between class groupings. At subsequent levels in the tree, the classification is refined until all classes are separated. Preliminary results indicate that using this technique perfect classification is achievable under the conditions in Case 2. This gain is due primarily to the increased flexibility allowed in the design of the features and decision boundaries for a particular class. For example, in early stages, the classifier might chose to group like classes together, expending effort accurately disambiguating classes that easily differentiable. This allows more difficult classes to be addressed only after all others have been classified.

Another direction being explored is the effect the choice of base representation has on overall classification performance. It is important to note that the kernel for optimal separation maximizes the time-frequency difference given the base representation. In this work, the Rihaczek ambiguity function has been used as the base representation. Preliminary results indicate that *classifier performance is reduced* when the Wigner-Ville ambiguity function is used in place of the Rihaczek ambiguity function.

6. DISCUSSION

The class-dependent methodology has been applied to the problem of radar transmitter identification. The goal of this methodology is to produce kernels (and thus, TFRs) which will enable accurate classification without explicitly defining, *a priori*, the underlying features that differentiate individual classes. Several modifications to our earlier work were shown to be necessary to accurately classify individual radar signals. Using this augmented approach it has been shown that 100% correct classification can be achieved on this data set, provided exact time registration of the data is available. When this information is unavailable, a performance loss on the order of 3% was observed in the overall recognition performance.

A final point that needs to be considered for the future is the issue of time registration. We have proposed a possible solution to overcome the degradation in performance due to misalignment of the training and testing waveforms. This solution incorporates our class-dependent methodology within the framework of classification trees. Initial results appear promising, however, further work is required before any conclusions can be drawn.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Office of Naval Research (N00014-97-1-0082). The authors would like to thank Dr. Victor Chen (Naval Research Laboratory) and James Droppo (University of Washington) for their assistance in this work.

REFERENCES

1. J. McLaughlin, J. Droppo, and L. Atlas, "Class-dependent time-frequency distributions via operator theory," *Proceedings of the 1997 IEEE ICASSP*, **3**, 1997.
2. L. Atlas, J. Droppo, and J. McLaughlin, "Optimizing time-frequency distributions via operator theory," *Proceedings of the SPIE – The International Society for Optical Engineering*, **3162**, 1997.
3. J. McLaughlin, and L. Atlas, "Applications of operator theory to time-frequency analysis and classification," to appear in *IEEE Transactions on Signal Processing*, 1998.
4. V. C. Chen, "Time-frequency/time-scale analysis for radar applications," *http://airborne.nrl.navy.mil/~vchen/tftsa.html*, current as of June 15, 1998.
5. D. H. Kil, and F. B. Shin, *Pattern Recognition and Prediction with Applications to Signal Characterization*, AIP Press, Woodbury, New York, 1996.
6. L. L. Scharf, *Statistical Signal Processing*, Addison Wesley, Reading Massachusetts, 1991.
7. L. Breiman, J. H. Freidman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth International Group, Belmont, California, 1984.